Group model building evaluation in single cases: a method to assess changes in mental models

Brigit Fokkinga
Inge Bleijenbergh
Jac Vennix[1]

Radboud University Nijmegen
Institute of Management Research
PO Box 9108
6500 HK Nijmegen
The Netherlands
Tel ++31 24 15608
B.Fokkinga@fm.ru.nl
I.Bleijenbergh@fm.ru.nl
J.Vennix@fm.ru.nl

Abstract

Group model building is supposed to support the development of shared mental models. This paper focuses on the development of a method for the assessment of changes of mental models in applied settings. A systematic approach is needed to be able to compare applied cases. The development of a measurement instrument opens up the possibility of comparative case-analysis. This instrument should be time-efficient and easy to apply for participants. We test our instrument in a particular case.

Key words: Group Model Building, measuring effectiveness, reconstruction of mental models, qualitative methodology, evaluation of single cases, comparative case-analysis.

[1] With thanks to dr. Marloes van Engen of Tilburg University, coordinator of the research project at Tilburg University, for supporting the collection of the written questionnaires.

1. Introduction

Outcomes of system dynamics (SD) interventions are difficult to test, even though case descriptions and effectiveness studies have been published almost since the inception of the field. Rouwette, Vennix and van Mullekom (2002) conducted a meta-analysis of 107 cases and show that outcomes of SD interventions are assessed very differently. This leads to a huge number of largely incomparable case descriptions. The authors strongly advocate the need for a more systematic approach towards assessment of SD interventions.
They also show that the variability of SD interventions is large. The variability relates to differences in the context of interventions: some interventions are performed in the context of public policy, while others cover issues within private organizations or non-governmental organizations. With regard to the content, SD interventions cover many different fields, ranging from defense, biology, and medicine, to environmental issues, social policy, and health care. The aim of the intervention can vary from building a general model to generate consensus about the problem at stake to building a detailed quantitative model to test concrete policy options. Also the role of the facilitator leading the intervention may vary from an expert role to the role of a process facilitator (Schein, 1987; Vennix, 1996) and interventions can be performed in a commercial or in a scientific setting, which may result in different ways of reporting regarding the methods of investigation.

Evaluation of Group Model Building, as a specific form of SD interventions in organizations, focusses on different variables, like changes in attitudes of participants and phase of the process that is supported. Moreover different methodologies are used to perform the evaluation, like questionnaires and observations. Rouwette (2003) compares five GMB cases from applied settings with regard to the attitudes towards the problem and commitment of participants before and after the intervention. By using individual pre- and posttest questionnaires he gathers data for each case. On the basis of quantitive analyses he finds that GMB can change attitudes of participants. He also finds that under certain conditions GMB can lead to an increase of commitment of participants.
Dwyer and Stave (2008) compare two cases from applied projects on urban growth in Las Vegas over a period of two years. One case follows the process and outcomes of traditional group facilitation, the other case encompasses a group model building approach. Dwyer and Stave collect data from minutes, written documents and direct observations of both interventions. By systematically comparing the cases on the basis of qualitative methods, they discovered that GMB supports stakeholders to pay attention to the causes of a problem more than traditional facilitation. In contrast, the traditional group facilitation process mainly induces participants to talk about alternatives.
McCardle-Keurentjes (2008) takes an experimental approach to asses the effectiveness of Group Model Building. Her experiment involves groups of students who perform the same problem-solving task. They are either in a GMB-group with a facilitator or in a group that performs a meeting as usual (no facilitation). She gathers data by several methods: individual pre- and posttest questionnaires, group questionnaires and video registration. By comparing results of groups with different kinds of decision-making support, she is able to assess the results of Group Model Building.

Of course an experimental or quasi experimental research design is the most suitable way to assess if Group Model Building produces changes in mental models. However, Group Model Building is often performed in the context of a single organization, where no control group can be found. If we want to compare outcomes from a single GMB-case we need another approach where mental models before and after the intervention can be measured.

Because in applied settings a control group is missing, we need a measurement instrument that allows for pre- and posttest measurement, like Rouwette (2003) did. However, we need to measure mental models rather than attitudes and commitment and we need a measurement instrument that can be applied in all settings and is time efficient. O' Connor, Johnson and Kahlil (2004) have developed a way to measure mental models in teams in applied settings, by instructing participants to create a causal loop diagram of their mental model of the problem at stake. This, however, is very time-consuming and asks for a high commitment of the participants to the assessment of the research.

Because respondents are often stakeholders with a limited time-schedule, a measurement-instrument is needed where data-collection allows for reconstructing mental models and also takes little time from the respondents. We argue that with such an instrument it is possible to analyze and report single cases in a more systematic way, which would enable secondary comparative case-analysis and so contribute to the generalization of knowledge from individual cases. We will focus on one of the supposed effects of Group Model Building, i.e. the development of a shared mental model of a particular problem within teams.

Before explaining our method, we need to define the concept of mental model. The definition of a mental model is essentially contested (Richardson, Andersen, Maxwell & Stewart, 1994). We will use the definition developed by Doyle and Ford (1998) on the basis of an extensive review in literature from System Dynamics and cognitive sciences. The definition was amended by Lane (1999) and resulted in the following definition that has been adopted by several authors (see for instance Rouwette, 2003):

"A mental model of a dynamic system is a relatively enduring and accessible, but limited, internal conceptual representation of an external system (historical, existing or projected) whose structure is analogous to the perceived structure of that system." (Doyle & Ford, 1999: 414).

In this paper we will use the definition by Doyle&Ford to formulate an operational definition:

> A mental model is a conceptual representation of a social problem that can be externalized in the form of a causal loop diagram.

What we need is a method that is able to reconstruct individual's mental models in causal loop diagrams and determine the degree to which they are shared within a group. On the basis of such a method, comparison of the changes in mental models of participants both within groups and over time will be enabled.

The contribution we want to make is twofold. First we want to contribute to the methodological development of a measurement instrument for assessing changes in mental models by group model building in single cases. We do so by proposing that such a measurement instrument should involve detecting changes in mental models on both individual and group level. Second, we want to focus on the measurement of single cases and support the analysis of the outcomes of these particular cases. In this way we aim to go beyond the regular approach towards single case studies, which is mainly descriptive in character as Rouwette et al. (2002) have shown, and open up the possibility for comparing outcomes of single cases in a meta analysis.

This study illustrates our first efforts to develop and test a measurement instrument for the evaluation of changes in mental models in single case-evaluations. The instrument will be based on the reconstruction of individual mental models and a comparison between these models before and after the intervention, and by comparing these mental models within and between groups. We pose two research questions:

How to measure changes in mental models in GMB interventions in applied settings?

What are the first results with this form of measuring in a concrete case and what can be learned from it?

2. Method

Rouwette (2003) discusses the validity of his measurements on participants of GMB-cases and finds that self-reporting of individuals about their own learning-process is highly invalid. Classical research from Nisbett and Wilson (as referred to in Rouwette, Bleijenbergh, Peters & van Mullekom, 2008) shows that people have little insight whether they have learned something, and if they have, what causes their learning.

Alliger and Janak (1989) provide empirical evidence for the value of individuals' appreciation of an intervention. It turns out that appreciation has no correlation with outcomes of the intervention, neither at an individual level nor at an organizational level. So, how can we capture changes in individuals' mental models?

We have chosen for an approach to reconstruct individual mental models before and after the intervention. In this way it is possible to make different comparisons, between individual pre- en posttest models, between individual models within groups and between individual models and a group model. This approach has been developed by Vennix (1990) and refined by Verburgh (1994), and is based on the research by Axelrod (1976) on the role of mental models in problem solving in policy issues[2]. Vennix and Verburgh reconstructed mental models into causal loop diagrams in an experimental setting, by performing content analysis on policy texts the participants had to write.

We want to test this method in a small quasi-experimental design without randomization.

We used qualitative content analysis that preserves the relationships between concepts in a person's mental model. With help of written questionnaires, respondents were asked to answer the following four questions before and after the group model building process:

1. Please write down the central problem at stake.
2. Please describe the causes of the problem
3. Please describe the effects of the problem
4. Please mention three options to solve the problem.

[2]Axelrod uses the term cognitive maps for the reconstruction of mental models. Lane (1999) warns for confusing this with the method of cognitive mapping developed by Eden (1988) and Eden &Ackerman (2004a, 2004b). We decided not to use the term cognitive map but refer to the reconstruction of a mental model into a causal loop diagram.

The written answers were analyzed and coded by two of the researchers and converted into causal loop diagrams. As we will show in the results, these causal loop diagrams were compared on an individual and group basis.

3. Case description

Our ideal would be to use a GMB-case with real life decision-makers, but in this paper we use a case-study from the Netherlands we were involved in, to apply and test our evaluation of a GMB-case. The Group Model Building was performed in the context of a large scale qualitative study towards the under-representation of women in higher academic positions at a Dutch University (Van Engen, Bleijenbergh & Pauwe, 2008). The study was initiated by the executive board of the university in who's view this under representation is a serious problem. A research team of ten people conducted 44 semi-structured interviews, followed by a series of ten focus groups. The team used system dynamics model building to support the analysis of the enormous amount of material and the formulation of policy recommendations (Bleijenbergh, Blonk, Schulte & van Engen, 2008). Five of the ten researchers, involving three senior researchers and two junior researchers, participated in the Group Model Building process. An independent SD facilitator and SD modeler from another university supported the process. The other five researchers in the team agreed to function as a control group. This group involved three senior researchers and two junior researchers. Since we miss a post-test-measurement of one of the senior researchers, also the pre test measurement of this person is left out of the material. The control group is now limited to four persons.

To measure potential changes in the mental models of the members of the research group, we used a written questionnaire with the open ended questions that were discussed above. Before we began with the GMB sessions, we asked all ten members of the research team to write down their ideas on this particular problem, the causes of the problem, the effects and possible solutions. This didn't make a difference for the absolute figures on feedback loops and content of variables. At that moment respondents had knowledge of the problem through their involvement in interviewing academics and content analysis of policy documents. Because of the huge amount of data, knowledge of the material was however dispersed over the different group members. After the written questionnaires were completed, the five members of the GMB group started the sessions, while five others were not involved (the control group). The control group had access to the causal loop diagram after it was finished, since it became part of the research report. So both groups had knowledge on the problem at stake and both groups were able to view the content of the causal loop diagram after it was finished. But only the members of the GMB group had been involved in building the model themselves.

Coding process

In total we collected 18 written questionnaires. The coding process was open, close to the data. We transformed concepts into variables that can increase or decrease, and are neutral and positively formulated. We also identified relationships between variables. The variables and relationships together made a causal loop diagram that we recorded with the program Vensim.

We started out with an exploratory coding of two respondents (four questionnaires) with two researchers to develop our inter-rater-reliability. One researcher stayed very close to the way respondents formulated their answers. The other researcher turned out to make more interpretations in the sense of identifying relations that can be supposed to be common knowledge in the work field of the respondents. The differences in coding between the two researchers were discussed, made explicit and agreed upon. This agreement was written down in a memo. Upon this agreement one researcher coded the other fourteen questionnaires. The resulting models were checked by the other researcher. The differences in coding between the two researchers decreased over time, so we assume that the reliability increased.

We will now give an example of the causal loop diagrams of a respondent from our control-group. The first diagram is the pretest model, the second one the posttest model from the same person.

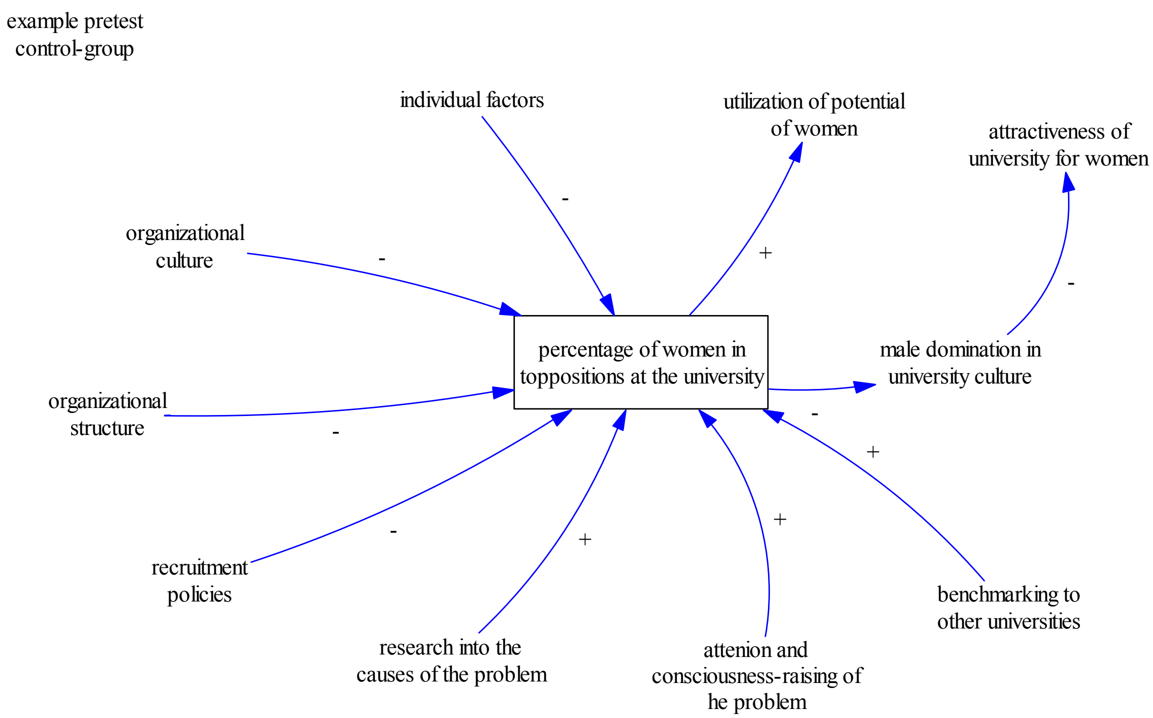

Figure 1: Example of pretest causal loop diagram

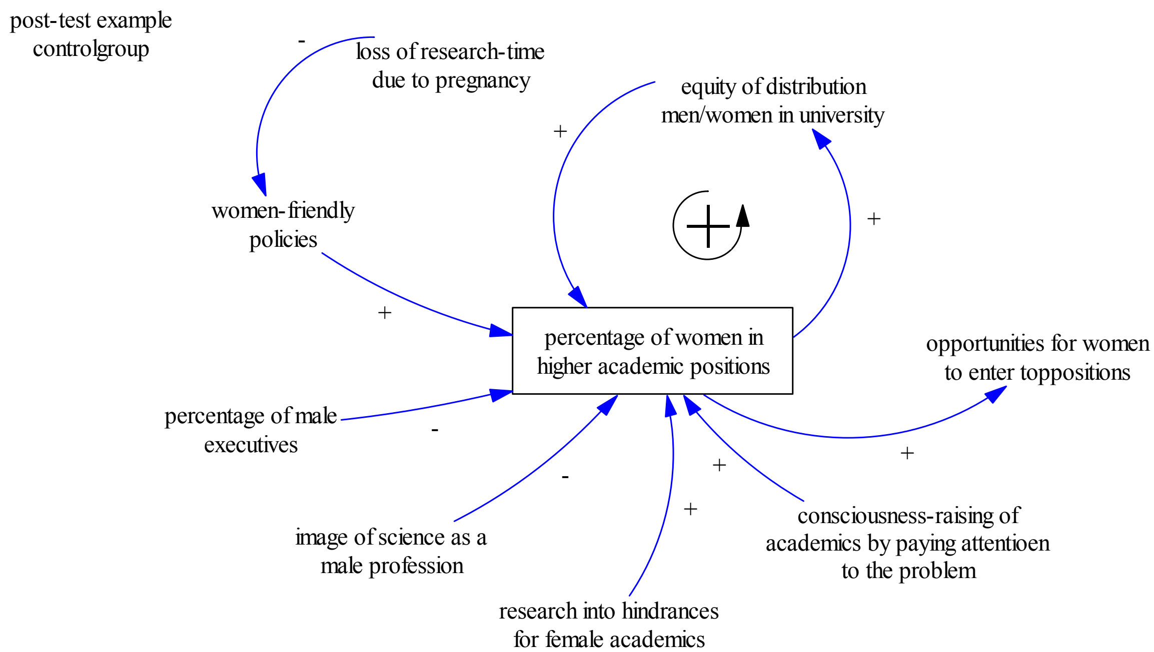

Figure 2: Example of posttest causal loop diagram from the same person

We encountered some problems with the aggregation level of variables. Some respondents are very general in the concepts they use while another respondent can be very specific and detailed. We solved this by adding names for clusters of variables in the diagram to show that people can vary extensively in the amount of variables they mention while still talking about the same concepts. We also decided to use neutral formulations, unless this would change the meaning of the respondent's argument so much that it would hinder recognition.

4. Results

What were the results of application of this method to the Group Model Building case? We will discuss the results by contrasting the characteristics of the research group with the characteristics of the control group. We will first discuss the number of variables the respondents connect to the problem. After that we discuss the content of the concepts (variables) respondents connect to the problem. Finally we will discuss the insight participants have in feedback processes related to the problem. The first effort we made was comparing the amount of variables respondents related to the problem at stake, the under-representation of female academics in top positions at a Dutch University.

Number of variables

As can be seen in table 1, the members of the GMB on average already reported more variables connected to the problem (16) than the respondents in the control group (11). So there were considerable differences between these two groups with regard to their knowledge of the problem at the start of the measurement. The fact that two of the senior researchers in the GMB group were gender experts, while only one in the control group was a gender expert, may explain this difference. However, these differences increased after the Group Model Building took place. Respondents involved in the Group Model Building on average reported 25 variables related to the problem, while the members of the control group reported on average 14 variables in the posttest after just reading about the model[3]. So having access to the causal loop diagram may have increased the knowledge on the number of variables slightly (on average from 11 to 14 variables), but building the model themselves seems to have had a much bigger impact on the knowledge of respondents on the amount of variables related to the problem at stake (from 16 to 25).

	GMB group		Control group	
Individual causal loop diagram	Pre-test	Post-test	Pre-test	Post-test
Amount of variables (mean)	16	25	11	14
Amount of variables similar to Group Model (mean)	0,8	2	1,25	1
Amount of feedbackloops (mean)	1	1	0.5	0,75

Table 1: Pre- and posttest causal loop diagrams of the GMB-group versus the control group

(The mean amounts of variables, variables similar to the Group Model, and feedback loops are the individual averages in a group.)

[3] Some will argue that reading the model instead of being involved in building of the model are two different stimuli that require two different research groups. We choose to focus on the process of group model building as one stimulus, since both groups were able to read the model afterwards and so do not differ with regard to that issue.

Content of variables

We found no substantial differences between the members of the GMB group and the members of the control group with regard to the content of concepts. In fact, content analysis of the written questionnaires showed that respondents mentioned very different variables. We measured the number of variables in the pretest material that corresponded to the variables in the causal loop diagram. In the GMB group we found an average of 0,8 corresponding variables, where we found 1n average of 1,25 in the control group. So, although the members of the GMB group on average had more knowledge (related more variables to the problem) before the GMB sessions started, this didn't result in a substantially larger number of corresponding variables in the later causal model than the control group. The influence seems to work the other way around: building the model contributed to the knowledge of the participants. After the GMB sessions took place, members of the GMB group showed an average of 2 variables that were similar to the GMB model, while the members of the control group only showed an average of one variable that was similar. So the model building effort supported the knowledge of the participants to be better tuned to each other.

Feedback loops

Did the Group Model Building also support learning about feedback processes related to the problem at stake? In other words: what were the differences in the amount of feedback loops respondents reported before and after the GMB sessions? The most striking result is the low presence of references to feedback processes in both the pretest and posttest material. The written questionnaires seemed to have stimulated respondents to report extensively on variables connected to the problem. But only seldom we found that respondents explicitly mentioned that a certain effect caused to further increase or decrease of the problem. Moreover, the amount of feedback loops reported didn't increase after the model building. Neither participants in the group model sessions (the GMB group), neither the control group showed an increase in the amount of reported feedback loops after the sessions had been finished.

We used the construction of causal loop diagrams to support content analysis of written questionnaires of participants to a group model building session and of a control group with comparable knowledge that was not involved in the model building. On the basis of this method we concluded that the amount of variables that respondents connected to the problem increased after a causal model of the problem was produced. However, the amount of variables increased much more for respondents that had been participating in the model building themselves than for the other respondents. Moreover, the content of the variables changed for the participants to the GMB sessions, while they remained the same for the control group. No direct effect of the causal modeling could be found for the amount of feedback loops that were reported on the issue.

5. Conclusion

To conclude, how to measure changes in mental models in GMB interventions in applied settings? We need a method that is able to measure changes in the mental models of both individuals and groups. Only by comparing the mental models before and after the group model building, we are able to determine if the intervention contributed to the *sharedness* of mental models. Moreover, the measuring should be relatively easy to apply. Finally, measuring should give information on both the knowledge of participants related to the content and their insight in feedback processes of the problem at stake.

In this paper we presented a measurement that meets these criteria. By comparing the mental models before and after the GMB such an instrument is able to measure changes in mental models at an individual level. Moreover, by comparing the mental models within and between groups such an instrument is able to measure changes in the *sharedness* of mental models on group level. We focused on measuring whether the variables of the causal loop diagram were articulated more often after the GMB took place? By using a rather simple written questionnaire, which takes about ten minutes to complete, we were able to measure the mental models of participants to a group model building sessions and of a control group before and after the GMB. By using causal loop diagrams we were able to reconstruct the mental models of the participants.

What are the first results of measuring a GMB intervention with this form of measuring in a concrete case and what can be learned from it?

In this case it looks as if the production of a causal loop diagram increased the knowledge on a problem with regard to the number of variables connected to the problem. However, this increase was stronger for respondents that participated in the model building than for the respondents that were only able to read it afterwards. Moreover, it looked as if the content of the variables changed for the participants to the GMB sessions, while they remained the same for the control group. We did not find a direct effect of the causal modeling on the amount of feedback loops respondents reported. Neither the group participating in the model building, nor the control group, reported more feedback loops after the causal diagram was produced and reported. So, our instrument seems to look useful.

6. Discussion

Our results may indicate that there is a considerable difference between the effect of studying a causal loop diagram and the effect of being involved in building the causal model in a Group Model Building setting. The efforts of building a causal model in a group seems to increase the knowledge of the participants regarding the number of variables related to a particular social problem more than passively reading the model. This supports the idea that the strongest learning effect comes from being actively involved in formulating knowledge rather than from passively reading or listening to knowledge.

These results support a plea for actively involving practitioners in the building of a SD model on a particular problem rather than building the model *for* them. However, interestingly enough, our research didn't show an increase in the knowledge of feedback processes related to the problem. At least these feedback processes couldn't be reconstructed from the written analysis respondents made of the particular problem at stake. Maybe the reconstruction of CLD loops through content analysis may explain why we didn't find an increase in feedback loops. It is possible that feedback processes can more easily be communicated via visualization than by verbalization and that we were simply not able to distract them from the written material.

It is also possible that the knowledge on feedback processes simply didn't increase after a causal loop diagram on the problem was produced. Doyle, Radzicki and Scott Trees (1996) refer to research into the storage of newly learned mental models in the memory. Apparently, being actively involved in the building of a model seems not to be enough for storage in the long term memory; the old mental model kind of competes with the new model. Maybe some kind of reinforcement is needed for internalization of the new model. We plan to collect more information on this issue.

In this paper we presented the first step in the development of a method to assess single cases. We tested our measurement instrument on a team of researchers, of which five team-members were part of the Group Model Building group and four team members formed a control group. The next step in the development of the method would be the application in a real-life setting, in a case with a team of decision-makers. We will also use this case to perform an in depth analysis of the content of the causal loop diagrams.

References

Alliger, G.M., and Janak, A., (1989)
Kirkpatrick's levels of training criteria: thirty years later. *Personnel Psychology* 42: 331-342.

Axelrod, R., (ed.) (1976)
Structure of decision: the cognitive maps of political elites, Princeton (NJ).

Bleijenbergh, I., Blonk, E., Schulte, L. and Engen, M. van (2008)
Women's Careers at a Dutch university: Model building as a method for knowledge elicitation in gender analysis. *Proceedings of the 26th Conference of the System Dynamics Society*, Athens, Greece.
www.systemdynamics.org/conferences/2008/proceed/papers/BLEIJ307.pdf

Doyle, J.K. and Ford, D.N. (1998)
Mental models concepts for system dynamics research. *System Dynamics Review,* 14, 1, 3-29.

Doyle, J.K. and Ford, D.N. (1999)
Mental models concepts revisited: some clarifications to and a reply to Lane. *System Dynamics Review*, 15, 4, 411-415.

Doyle, J.K., Radzicki, M.J. and Trees, W.S. (1996)
Measuring the Effect of Systems Thinking Interventions on Mental Models. *Proceedings of the 1996 International System Dynamics Conference:* 129-132.

Dwyer, M. and Stave, K, (2008)
Group Model Building Wins: The Results of a Comparative Analysis. *Proceedings of the 26th International Conference of the System Dynamics Society.* Athens,Greece.
www.systemdynamics.org/conferences/2008/proceed/papers/STAVE364.pdf

Eden, C. and Ackerman, F. (2004a)
Cognitive mapping expert views for policy analysis in the public sector. *European Journal of Operational Research*, 152: 615-630.

Eden, C. and Ackerman, F. (2004b)
Analyzing cognitive maps to help structure issues or problems. *European Journal of Operational Research*, 159: 673-686.
Eden, C. (1988)
Cognitive mapping. *European Journal of Operational Research,* 36: 1-13.
Lane, D.C. (1999)
Friendly amendment: A commentary on Doyle and Ford's proposed re-definition of 'mental model'. *System Dynamics Review,* 15, 2, 185-194.
McCardle-Keurentjes, M.H.F. (2008)
Effectiveness of Group Model Building in discovering hidden profiles in strategic decision-making. *Proceedings of the 26th International Conference of the System Dynamics Society.* Athens, Greece.
www.systemdynamics.org/conferences/2008/proceed/papers/MCCAR357.pdf
Richardson, G.P., Andersen, D.F., Maxwell, T.A. and Stewart, T.R. (1994)
Foundations of Mental Model Research. *Proceedings of the 1994 International System Dynamics Conference.* Stirling, Scotland.
www.systemdynamics.org/conferences/1994/proceed/papers_vol_2/richa181.pdf
Rouwette, E.A.J.A (2003)
Group model building as mutual persuasion. PhD dissertation Nijmegen University: Wolf Legal Publishers.
Rouwette, E., Vennix, J. and Mullekom, T. van, (2002)
Group model building effectiveness: a review of assessment studies. *System Dynamics Review,* 18(1): 5-45.
Rouwette, E.A.J.A., Bleijenbergh, I.L., Peter, V.A.M., Mullekom, T. van (2008)
Consensus bouwen in een probleemwijk. Modelbouw als interventiemethode bij het bestrijden van overlast. *Tijdschrift M&O, 6, 5-21.*
Schein, E.H. (1987)
Process Consultation. Volume 2: Lessons for Managers and Consultants. Reading (MA): Addison-Wesley
Vennix, J.A.M. (1996).
Group Model-Building: facilitating team learning using System Dynamics.
Chichester: John Wiley and Sons.
Vennix, J.A.M. (1990)
Mental models and computer models. Design and evaluation of a computer-beased learning environment for policy-making. PhD dissertation, Nijmegen University.
Verburgh, L.D. (1994)
Participative Policy Modelling: Applied to Health Care Insurance Industry. PhD dissertation Nijmegen University
Van Engen, M., Bleijenbergh, I. en Paauwe, J. (2008)
Vrouwen in hogere wetenschappelijke posities aan de Universiteit van Tilburg: Processen van instroom, doorstroom en uitstroom nader bekeken. Rapportage aan College van Bestuur en Decanen. Tilburg: Universiteit van Tilburg.
www.uvt.nl/medewerkers/vrouwen.html